

STIC-Biotech/ChemLib

78418

Paper 20 Attach.

From: Loeb, Bronwen
Sent: Tuesday, October 22, 2002 3:15 PM
To: STIC-Biotech/ChemLib
Subject: sequence search request 09/717,321

09/717,321
Rothberg et al
filed Nov. 20, 2000

Please search SEQ ID NO. 1 in all databases.

Thank you.

Bronwen Loeb, PhD

AU 1636
703-605-1197
CM1 11D-16
Mailbox 11E-12

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6B06 TEL. (703) 305-1954

RECEIVED
OCT 22 2002
STIC

Start 10/24 10
Fin 10/28 10

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 26, 2002, 03:02:20 ; Search time 1600 Seconds
(without alignments)
1037.578 Million cell updates/sec

Title: US-09-717-321a-1
Perfect score: 123
Sequence: 1 caattgaaagtttttc.....ctcacagaccacaaagtacc 123

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST.*
- 1: em_estba.*
 - 2: em_esthum.*
 - 3: em_estin.*
 - 4: em_estnu.*
 - 5: em_estov.*
 - 6: em_estpl.*
 - 7: em_estro.*
 - 8: em_htc.*
 - 9: gb_estl.*
 - 10: gb_est2.*
 - 11: gb_hc.*
 - 12: gb_gss.*
 - 13: em_gss_hum.*
 - 14: em_gss_inv.*
 - 15: em_gss_pla.*
 - 16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	122	99.2	277	10	BF420446
2	122	99.2	292	9	AI598992
3	122	99.2	316	9	AI412434
4	122	99.2	404	10	BF549054
5	122	99.2	405	9	BE104696
6	122	99.2	407	9	AI009011
7	122	99.2	429	9	AI059212
8	122	99.2	455	9	AI012196
9	122	99.2	463	10	BF548903
10	122	99.2	465	9	AR99757
11	122	99.2	477	9	AW520555
12	122	99.2	489	9	AA859879
13	122	99.2	493	10	BF554834
14	122	99.2	522	10	C06842
15	122	99.2	565	10	BI273986
16	122	99.2	615	10	BG380448
17	122	99.2	639	9	BE111691

18	122	99.2	687	10	BM389059
19	120.4	97.9	316	9	AI409795
20	120.4	97.9	557	10	BG665005
21	118.8	96.6	517	9	AA859940
22	118.8	96.6	553	9	AA799542
23	118.8	96.6	638	10	BI283790
24	117.2	95.3	294	9	AI236740
25	117.2	95.3	469	10	BI276611
26	115.6	94.0	431	9	AI410870
27	115.6	94.0	446	10	BI286743
28	112.4	91.4	256	9	AI234716
29	110.8	90.1	427	9	AI178082
30	107.6	87.5	225	9	BE136269
31	107.6	87.5	240	9	AW323013
32	107.6	87.5	283	9	AW122670
33	107.6	87.5	288	9	AV128839
34	107.6	87.5	294	10	W13019
35	107.6	87.5	298	9	AA940186
36	107.6	87.5	299	9	BB749463
37	107.6	87.5	309	9	AA265870
38	107.6	87.5	317	9	AA396854
39	107.6	87.5	340	9	AA388802
40	107.6	87.5	356	9	AA163447
41	107.6	87.5	363	9	BB686346
42	107.6	87.5	376	10	BM200395
43	107.6	87.5	378	9	AA461828
44	107.6	87.5	380	9	AA213220
45	107.6	87.5	386	9	BB685587

ALIGNMENTS

RESULT 1
BF420446
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF420446 277 bp mRNA linear EST 28-NOV-2000
UI-R-BJ2-bpw-e-11-0-UI-s1 UI-R-BJ2 Rattus norvegicus CDNA clone
UI-R-BJ2-bpw-e-11-0-UI 3', mRNA sequence.
BF420446
EST.
GI:11408475
Norway rat.
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 277)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized atrium at 16.5 dpc library cDNA Library Preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-Yes.

FEATURES
source
Location/Qualifiers
1..277
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"

```
/db_xref="taxon:10116"
/clone="UI-R-BJ2-bpw-e-11-0-UI"
/clone_lib="UI-R-BJ2"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-BJ2
library is a subtracted library derived from the following
tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc,
atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15
dpc, AV canal at 15 dpc. For a detailed description of
the library from which this clone was derived, please
visit our web site at ratest.eng.uiowa.edu. The
subtraction has been previously described in (Bonaldi,
Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=UI-R-BJ2
TAG_ISSUE=atrium at 16.5 dpc
TAG_SEQ=GATTC
BASE COUNT      85 a      69 g      65 t
ORIGIN

Query Match      99.2%; Score 122; DB 10; Length 277;
Best Local Similarity 100.0%; Pred. No. 3e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTCTTCCAGTGAG 60
DB 66 CAATTGAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTCTTCCAGTGAG 125
QY 61 TTAGTTGTACAGACGGCGTTAGCAGTGTGACAGACCTTCACAGACCCAAAGGT 120
DB 126 TTAGTTGTACAGACGGCGTTAGCAGTGTGACAGACCTTCACAGACCCAAAGGT 185
QY 121 AC 122
DB 186 AC 187

RESULT 2
LOCUS      A1598992      292 bp      mRNA      linear      EST 21-APR-1999
DEFINITION      E2250695 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
ACCESSION      A1598992
VERSION      A1598992
KEYWORDS      EST.
SOURCE      Rattus sp.
ORGANISM      Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 292)
Lee,N.H., Glodex,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1. 292
/organism="Rattus sp."
/db_xref="taxon:10116"
/clone="REM05"
/clone_lib="Normalized rat embryo, Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/notes="Vector: pRT3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      89 a      60 c      70 g      73 t
ORIGIN

Query Match      99.2%; Score 122; DB 9; Length 316;
Best Local Similarity 100.0%; Pred. No. 3.1e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTCTTCCAGTGAG 60
DB 97 CAATTGAAAGTTTGTCTAGTGTGCGAAAGGCCCAACACTGTCTTCCAGTGAG 156
QY 61 TTAGTTGTACAGACGGCGTTAGCAGTGTGACAGACCTTCACAGACCCAAAGGT 120
DB 157 TTAGTTGTACAGACGGCGTTAGCAGTGTGACAGACCTTCACAGACCCAAAGGT 216
QY 121 AC 122
DB 217 AC 218

RESULT 4
LOCUS      A1412434      316 bp      mRNA      linear      EST 09-FEB-1999
DEFINITION      EST240733 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
ACCESSION      A1412434
VERSION      A1412434.1 GI:4255938
KEYWORDS      EST.
SOURCE      Rattus sp.
ORGANISM      Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 316)
Lee,N.H., Glodex,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,M.D.
Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
Gene Index
Unpublished (1998)
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers
1. 316
/organism="Rattus sp."
/db_xref="taxon:10116"
/clone="RBD065"
/clone_lib="Normalized rat brain, Bento Soares"
/notes="Organ: brain; Vector: pRT3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      98 a      62 c      76 g      80 t
ORIGIN
```

```

BF549054/c
LOCUS      BF549054      404 bp mRNA linear EST 11-DEC-2000
DEFINITION UI-R-A0-ag-b-01-0-UI.r1 UI-R-A0 Rattus norvegicus cDNA clone
ACCESSION  BF549054
VERSION     BF549054.1 GI:11640209
KEYWORDS   EST.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 404)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
JOURNAL   discovery
MEDLINE   97044477
COMMENT   Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msobres@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            The following repetitive elements were found in this cDNA sequence:
            1-22, >AT-richLow_complexity
            Seq primer: M13 Forward
            POLYA=yes.

FEATURES             Location/Qualifiers
     source           1..404
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-A0-ag-b-01-0-UI"
                     /clone_lib="UI-R-A0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT73D-Pac (Pharmacia) with a modified
                     polylinker; Site.1: Not 1; Site.2: Eco RI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from rat placenta, adult lung, brain
                     liver, kidney, heart, spleen, ovary, and muscle. The tag
                     is a string of 3-5 nucleotides present between the Not I
                     site and the oligo-dT track which allows identification of
                     the library of origin of a clone within the mixture."

BASE COUNT      92 a      98 c      83 g      131 t
ORIGIN
Query Match      99.2%; Score 122; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 3.3e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 CAATTGAAAAGTTTCTCTAGTGGTCGAAAGCCCAACACTGTGTCTTCCAGTGAG 60
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  368 CAATGAAAAGAGTTTCTCTAGTGGTCGAAAGCCCAACACTGTGTCTTCCAGTGAG 309
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY  61 TTAGTGTGTACAGAACGCGTTAGCACTAGCGCTTGACAGAACTCTACAGACCCCAAGGT 120
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  308 TTAGTGTGTACAGAACGCGTTAGCACTAGCGCTTGACAGAACTCTACAGACCCCAAGGT 249
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY  121 AC 122
    ||
Db  248 AC 247

RESULT 5
LOCUS      BE104696      405 bp mRNA linear EST 13-JUN-2000
DEFINITION UI-R-BX0-ars-e-01-0-UI.s1 UI-R-BX0 Rattus norvegicus cDNA clone
ACCESSION  BE104696
VERSION     BE104696
KEYWORDS   EST.
SOURCE     BE104696.1 GI:8496796
ORGANISM   Rattus sp.

Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE  1 (bases 1 to 404)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
JOURNAL   discovery
MEDLINE   97044477
COMMENT   Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msobres@blue.weeg.uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward.

FEATURES             Location/Qualifiers
     source           1..404
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-A0-ag-b-01-0-UI"
                     /clone_lib="UI-R-A0"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT73D-Pac (Pharmacia) with a modified
                     polylinker; Site.1: Not 1; Site.2: Eco RI; This library
                     consists of a mixture of individually tagged normalized
                     libraries constructed from rat placenta, adult lung, brain
                     liver, kidney, heart, spleen, ovary, and muscle. The tag
                     is a string of 3-5 nucleotides present between the Not I
                     site and the oligo-dT track which allows identification of
                     the library of origin of a clone within the mixture."

BASE COUNT      92 a      98 c      83 g      131 t
ORIGIN
Query Match      99.2%; Score 122; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 3.3e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 CAATTGAAAAGTTTCTCTAGTGGTCGAAAGCCCAACACTGTGTCTTCCAGTGAG 60
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  368 CAATGAAAAGAGTTTCTCTAGTGGTCGAAAGCCCAACACTGTGTCTTCCAGTGAG 309
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY  61 TTAGTGTGTACAGAACGCGTTAGCACTAGCGCTTGACAGAACTCTACAGACCCCAAGGT 120
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  308 TTAGTGTGTACAGAACGCGTTAGCACTAGCGCTTGACAGAACTCTACAGACCCCAAGGT 249
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY  121 AC 122
    ||
Db  248 AC 247

RESULT 6
LOCUS      AI009011      407 bp mRNA linear EST 25-JAN-1999
DEFINITION EST203462 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
ACCESSION  AI009011
VERSION     AI009011.1 GI:3222843
KEYWORDS   EST.
SOURCE     Rattus sp.
ORGANISM   Rattus sp.

```

REFERENCE	1 (bases 1 to 429)
AUTHORS	Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL	Genome Res. 6 (9), 791-806 (1996)
MEDLINE	97044477
COMMENT	Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoares@blue.weeg.uiowa.edu The sequence tag present in the cDNA between the NotI site and the oligo-dr track served to identify it as a clone from the normalized

RESULT 8.	
AI012196	linear EST 15-JUN-1998
LOCUS	455 bp mRNA
DEFINITION	EST206647 Normalized rat placenta, Bento Soares Rattus sp. cDNA
ACCESSION	EST206647
VERSION	AI012196
KEYWORDS	clone RPLAT85 3' end, mRNA sequence.
SOURCE	AI012196.1 GI:3226028
ORGANISM	Rattus sp.
	Rattus sp.
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
REFERENCE	Rattus.
	1 (bases 1 to 455)

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A0-ag-b-08-0-UI"
/clone_lib="UI-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
, liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dT track which allows identification of
the library of origin of a clone within the mixture."
BASE COUNT      109 a   106 c   93 g   155 t
ORIGIN
      Query Match          99.2%;   Score 122;   DB 10;   Length 463;
      Best Local Similarity 100.0%;   Pred. No. 3.4e-30;
      Matches 122;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

QY      1  CAATTGAAAAAGCTTTGTTCTTACTGTGTCGAAGGCCCAACACTGTGTCTTCTTGCAGTGAG 60
      |||
Db      368  CAATTGAAAAAGCTTTGTTCTTACTGTGTCGAAGGCCCAACACTGTGTCTTCTTGCAGTGAG 309
      |||

QY      61  TTAGGTTGTACAGAACGGCGCTTAGCACTAGCGTTGACAGAACTTCACAGACCCAAAGGT 120
      |||
Db      308  TTAGGTTGTACAGAACGGCGCTTAGCACTAGCGTTGACAGAACTTCACAGACCCAAAGGT 249
      |||

QY      121 AC 122
      ||
Db      248 AC 247

RESULT 10
AA899757
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT
FEATURES
source

```

```

/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: NotI; Site 2: EcoRI; This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dr track which allows
identification of the library of origin of a clone within
the mixture."
BASE COUNT      155 a      99 g      118 t
ORIGIN

```

```

Query Match      99.2%; Score 122; DB 9; Length 465;
Best Local Similarity 100.0%; Pred. No. 3.4e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CAATTGAAAAAGTTTCTAGTGTCTGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 60
    |||||||
Db 112 CAATTGAAAAAGTTTCTAGTGTCTGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 171
    |||||||
61 TTAGTTGTACAGAACGGCGTTAGCAGTGTGACAGAACCTTCACAGACCCAAAGGT 120
    |||||||
Db 172 TTAGTTGTACAGAACGGCGTTAGCAGTGTGACAGAACCTTCACAGACCCAAAGGT 231
    |||||||
QY 121 AC 122
Db 232 AC 233

```

```

RESULT 11
AW520555
LOCUS
DEFINITION
UI-R-BJOp-afx-g-09-0-UI-s1 UI-R-BJOp Rattus norvegicus cDNA clone
UI-R-BJOp-afx-g-09-0-UI 3', mRNA sequence.
ACCESSION
AW520555
VERSION
AW520555.1 GI:7162933
KEYWORDS
EST.
SOURCE
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE
1 (bases 1 to 477)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL
MEDLINE
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dr track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dr track served to identify it as a clone from the
normalized AV canal at 15 dpc library cDNA library preparation:
M.B. Soares Lab Clone distribution: clones will be available
through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

```

```

FEATURES
SOURCE
Location/Qualifiers
1..477
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJOp-afx-g-09-0-UI"
/clone_lib="UI-R-BJOp"

```

```

/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJOp
library is a subtracted library derived from the UI-R-AAL,
UI-R-ABI, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AE1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 15 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the Not I site
and the oligo-dr track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG-LIB-UI-R-BJOp
TAG-RISSUE-AV canal at 15 dpc
TAG-SEQ-GAAGG*
BASE COUNT      157 a      103 g      124 t
ORIGIN

```

```

Query Match      99.2%; Score 122; DB 9; Length 477;
Best Local Similarity 100.0%; Pred. No. 3.5e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CAATTGAAAAAGTTTCTAGTGTCTGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 60
    |||||||
Db 112 CAATTGAAAAAGTTTCTAGTGTCTGAAAGGCCCAACACTGTGTTCTTCCAGTGAG 171
    |||||||
61 TTAGTTGTACAGAACGGCGTTAGCAGTGTGACAGAACCTTCACAGACCCAAAGGT 120
    |||||||
Db 172 TTAGTTGTACAGAACGGCGTTAGCAGTGTGACAGAACCTTCACAGACCCAAAGGT 231
    |||||||
QY 121 AC 122
Db 232 AC 233

```

```

RESULT 12
AA859879
LOCUS
DEFINITION
UI-R-E0-cc-c-06-0-UI-s1 UI-R-E0 Rattus norvegicus cDNA clone
UI-R-E0-cc-c-06-0-UI 3', similar to dbj|D25274|HUMPO2ST9 Human
randomly sequenced mRNA, mRNA sequence.
ACCESSION
AA859879
VERSION
AA859879.1 GI:4230416
KEYWORDS
EST.
SOURCE
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

REFERENCE
1 (bases 1 to 489)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL
MEDLINE
COMMENT
On Mar 10, 1998 this sequence version replaced g1:2949399.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

```

```

The sequence tag present in the cDNA between the NotI site and the
oligo-dr track served to identify it as a clone from the normalized
adult 18-Day-Embryo library. cDNA Library Preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through
Research Genetics This clone is also available through the
I.M.A.G.E. Consortium at LNL (info@image.llnl.gov). IMAGE
ID=1770064
Seq primer: M13 Forward

```

FEATURES
SOURCE

Best Local Similarity 100.0%; Pred. No. 3.5e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCTGAAGAGGCCCAACACTGTGTCTTCCCACTGAG 60
|||||
Db 69 CAATTGAAAAAGTTTGTCTAGTGTGCTGAAGAGGCCCAACACTGTGTCTTCCCACTGAG 128
|||||
QY 61 TTAGTTTGTACAGAACGCGCTTGTAGCTAGCTAGCTGTGACAGAACCTCACAGACCCAAAGGT 120
|||||
Db 129 TTAGTTTGTACAGAACGCGCTTGTAGCTAGCTAGCTGTGACAGAACCTCACAGACCCAAAGGT 188
|||||
QY 121 AC 122
||
Db 189 AC 190

RESULT 15

BI273986

LOCUS

BI273986 565 bp mRNA linear EST 18-JUL-2001
UI-R-CW0-bwb-c-12-0-UI-s1 UI-R-CW0 Rattus norvegicus cDNA clone

DEFINITION

UI-R-CW0-bwb-c-12-0-UI 3', mRNA sequence.

BI273986

VERSION

BI273986.1 GI:14884485

KEYWORDS

EST.

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

REFERENCE

1 (bases 1 to 565)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

CONTACT: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dr track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dr track served to identify it as a clone from the

normalized rat aorta pool library cDNA library Preparation: M.B.

Soares Lab Clone Distribution: clones will be available through

Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA-res.

Location/Qualifiers

1..565

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-CW0-bwb-c-12-0-UI"

/clone_lib="UI-R-CW0"

/dev_stage="ADULT"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CW0

library is a normalized library constructed from the

following rat aorta tissues: embryonic day 19, embryonic

day 21, adult day 1, adult day 12, adult day 75, adult day

200. For a detailed description of the library from which

this clone was derived, please visit our web site at

rateateng.uiowa.edu. The subtraction has been previously

described in (Bonaldo, Lennon and Soares, Genome Research

6:791-806, 1996)

TAG_LIB=UI-R-CW0

TAG_TISSUE=rat aorta pool

BASE COUNT 188 a 107 c 146 t 1 others
ORIGIN

Query Match 99.2%; Score 122; DB 10; Length 565;
Best Local Similarity 100.0%; Pred. No. 3.6e-30;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAAGTTTGTCTAGTGTGCTGAAGAGGCCCAACACTGTGTCTTCCCACTGAG 60
|||||
Db 113 CAATTGAAAAAGTTTGTCTAGTGTGCTGAAGAGGCCCAACACTGTGTCTTCCCACTGAG 172
|||||
QY 61 TTAGTTTGTACAGAACGCGCTTGTAGCTAGCTAGCTGTGACAGAACCTCACAGACCCAAAGGT 120
|||||
Db 173 TTAGTTTGTACAGAACGCGCTTGTAGCTAGCTAGCTGTGACAGAACCTCACAGACCCAAAGGT 232
|||||
QY 121 AC 122
||
Db 233 AC 234

Search completed: October 26, 2002, 04:58:31
Job time : 1604 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	31.4	25.5	605	4	US-09-385-982-483	Sequence 483, Appl
2	25.6	20.8	3033	2	US-08-462-080B-1	Sequence 1, Appli
3	25.6	20.8	3033	3	US-08-462-080-1	Sequence 1, Appli
4	25.6	20.8	3033	3	US-08-463-461-1	Sequence 1, Appli
5	25.2	20.5	49272	1	US-08-614-770A-1	Sequence 1, Appli
6	25	20.3	550	3	US-08-705-771-1	Sequence 1, Appli
7	25	20.3	786	5	PCF-US95-08295-1	Sequence 1, Appli
8	24.6	20.0	3505	1	US-08-452-083-1	Sequence 1, Appli
9	24.6	20.0	80595	4	US-09-078-234-3	Sequence 3, Appli
10	23.4	19.0	228	1	US-08-250-958-3	Sequence 3, Appli
11	23.4	19.0	231	5	PCF-US95-00605-2	Sequence 2, Appli
12	23.4	19.0	588	4	US-09-385-982-133	Sequence 133, App
13	23.4	19.0	741	3	US-07-927-391-25	Sequence 25, Appl
14	23.2	18.9	576	1	US-08-086-438B-11	Sequence 11, Appl
15	23.2	18.9	576	2	US-08-468-570-11	Sequence 11, Appl
16	23.2	18.9	576	2	US-08-290-665A-11	Sequence 11, Appl
17	23.2	18.9	576	5	PCF-US95-10398-11	Sequence 11, Appl
18	23.2	18.9	3216	2	US-08-828-007-1	Sequence 1, Appli
19	23.2	18.9	3385	1	US-08-405-392-1	Sequence 1, Appli
20	23.2	18.9	3385	3	US-08-487-691-1	Sequence 1, Appli
21	23.2	18.9	3385	3	US-08-666-221B-9	Sequence 3, Appli
22	23.2	18.9	3385	3	US-08-666-221B-9	Sequence 9, Appli
23	23.2	18.9	4134	2	US-08-817-090B-1	Sequence 1, Appli
24	23.2	18.9	4137	2	US-08-817-090B-3	Sequence 3, Appli
25	23.2	18.9	4322	2	US-08-537-342-1	Sequence 1, Appli
26	23.2	18.9	5162	2	US-08-916-917-13	Sequence 13, Appl
27	23.2	18.9	5162	3	US-09-225-170-13	Sequence 13, Appl

RESULT 4
US-08-463-461-1
; Sequence 1, Application US/08463461
; Patent No. 6103464

```

GENERAL INFORMATION:
APPLICANT: FOWLER, Timothy
APPLICANT: Barnett, Christopher C.
APPLICANT: Shoemaker, Sharon
TITLE OF INVENTION: Saccharification of Cellulose by Cloning
TITLE OF INVENTION: and Amplification of the Beta-glucosidase Gene of
TITLE OF INVENTION: Trichoderma reesei
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Genencor International, Inc.
STREET: 925 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: U.S.A.
ZIP: 94304-1013
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,461
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Christopher L. Stone
REGISTRATION NUMBER: 35,696
REFERENCE/DOCKET NUMBER: GC78D4
TELEPHONE: 415-846-7555
TELEFAX: 415-845-6504
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3033 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoderma reesei
FEATURE:
NAME/KEY: CDS
LOCATION: join(311..375, 446..2205, 2270..2679)
FEATURE:
NAME/KEY: intron
LOCATION: 376..445
FEATURE:
NAME/KEY: intron
LOCATION: 2206..2269
US-08-463-461-1

Query Match 20.8%; Score 25.6; DB 3; Length 3033;
Best Local Similarity 55.7%; Pred. No. 3.9; Mismatches 0; Gaps 0;
Matches 49; Conservative 0; Indels 39;

QY 12 ACTTTGTTCTAGTTCGAAAGGCCCAACACTGTCTTCTGCGAGTAGTGGTTGTAC 71
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 260 AGTTCTTCTACCACTCTTGACCAAGACCACTCTGTCAGCCCAATCAGAAATGGTTAC 319
QY 72 AGAAGCGGGTTAGCACTAGCGCTTGACA 99
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 320 CGAAGCAGCGTGGCTGGCACTTGCCA 347

RESULT 5
US-08-614-770A-1/C
; Sequence 1, Application US/08614770A
; Patent No. 5773267
; GENERAL INFORMATION:
; APPLICANT: WILLIAM R. JACOBS AND GRAHAM F. HATFULL
; TITLE OF INVENTION: D29 SHUTTLE PHASMIDS AND USES THEREOF

```

```

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESS: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
MEDIUM TYPE: DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,770A
FILING DATE: MARCH 7, 1996
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/402
TELEPHONE: (212) 286-0854 or 286-0082
TELEFAX: TWK 710-581-4766
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 49272
TYPE: NUCLEIC ACID
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: OLIGONUCLEOTIDE
DESCRIPTION: NO
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIOPHAGE
INDIVIDUAL ISOLATE: D29
US-08-614-770A-1

Query Match 20.5%; Score 25.2; DB 1; Length 49272;
Best Local Similarity 55.8%; Pred. No. 16;
Matches 48; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 22 AGTGTGGAAGGCCCAACACTGTCTTCTGCGAGTAGTGGTTGTACAGACGGCGT 81
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 47920 AGTGTCAACGTCCGACCTCCGAGATCGTCGAGAGCGTGAGCTGCCCTGTCGAAGA 47861
QY 82 TAGCACTAGCGCTTGACAGAACCTCA 107
||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 47860 CATCAGATCACTGGCGACACCGTCA 47835

RESULT 6
US-08-705-771-1
; Sequence 1, Application US/08705771
; Patent No. 6054289
; GENERAL INFORMATION:
; APPLICANT: Paul Moore, Reiner Gentz, Hongjin Ji,
; APPLICANT: Jian Ni and Jing-Shan Hu
; TITLE OF INVENTION: Human Genes, Sequences and
; TITLE OF INVENTION: Expression Products
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE

```

Query Match	20.3%	Score 25;	DB 5;	Length 786;
Best Local Similarity	53.6%;	Pred. No. 3.9;		
Matches 52;	Conservative 0;	Mismatches 45;	Indels 0;	Gaps 0;
22	AGTGTGCGAAGGCCCACTGTGTCCTTGCCAGTGAGTTAGTTGTACAGAACGCGGT	81		
260	AGTGGCCGAGAGAACCAAGGAGCAGGCCAACGCCGTGAGCAAGGCTGTGTGACGCGT	319		
82	TAGCACTACCGCTTGACAGAACCTCACAGACCCAAAG	118		
320	CAAACATKTTGGCCCAACCAAGACCGCTKGAGGAGCGCGAG	356		

RESULT 9
US-09-078-294 -3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294

Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 35 CCCAACACTGTGTTCTGCGAGTGTAGGTGTACAGAACGCGGTAGCACTAGCGCT 94
 III III III I III I III III III III III III III III III
 Db 476 CCCGCCACCATGTCACAGACAGCTTGTGGGATTCGGAGTAACTGGATAGCACTAGCGCT 417

RESULT 15

US-08-468-570-11/c
 ; Sequence 11, Application US/08468570
 ; Patent No. 5871962
 ; GENERAL INFORMATION:
 ; APPLICANT: BURKH, J., MILLER, R.H. AND
 ; APPLICANT: PORCELL, R.H.
 ; TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
 ; TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 GENE
 ; TITLE OF INVENTION: OF 51 ISOLATES OF HEPATITIS C AND THE USE
 ; TITLE OF INVENTION: OF REAGENTS DERIVED FROM THESE SEQUENCES IN
 ; TITLE OF INVENTION: DIAGNOSTIC METHODS AND VACCINES
 ; NUMBER OF SEQUENCES: 159

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA
 ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERFECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/468,570
 FILING DATE: 6-JUN-1995
 CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/086,428
 FILING DATE: 29-JUN-1993
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: RICHARD W. BORK
 REGISTRATION NUMBER: 36,459
 REFERENCE/DOCKET NUMBER: 2026-4070US1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 758-4800
 TELEFAX: (212) 751-6849
 TELEX: 421792

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 576 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 ORIGINAL SOURCE:
 ORGANISM: homosapiens
 INDIVIDUAL ISOLATE: DK1
 US-08-468-570-11

Query Match 18.9%; Score 23.2; DB 2; Length 576;

Best Local Similarity 61.7%; Pred. No. 16;

Matches 37; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 35 CCCAACACTGTGTTCTGCGAGTGTAGGTGTACAGAACGCGGTAGCACTAGCGCT 94
 III III III I III I III III III III III III III III III
 Db 476 CCCGCCACCATGTCACAGACAGCTTGTGGGATTCGGAGTAACTGGATAGCACTAGCGCT 417

Search completed: October 26, 2002, 04:59:35

Job time : 54 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 26, 2002, 02:49:04 ; Search time 180 Seconds
(without alignments)
1173.225 Million cell updates/sec

Title: US-09-717-321a-1

Perfect score: 123

Sequence: 1 caattgaaaaaattgttc.....ctcacagaccacaaagtacc 123

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 347872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123	100.0	123	22	AAH22395 Human rac1 genomic
2	123	100.0	348	22	AAH22399 Human rac1 gene re
3	123	100.0	1266	22	AAH22396 Human rac1 contig
4	106.6	86.7	422	21	AAH99694 Mouse Rab2 nucleot
5	96.6	78.5	447	21	AAH99693 Mouse Exon103 nucle
6	79.6	64.7	352	22	AAH22400 Human rac1 gene re
7	79.6	64.7	1318	23	AAH55569 DNA encoding novel
8	79.6	64.7	2051	21	AAH18236 Lung cancer associ
9	38.2	31.1	3740	22	AAH07296 Human reproductive

10	31.4	25.5	605	21	AAH16478 Human colon cancer
11	27.4	22.3	1089	21	AAC50142 Arabidopsis thalia
12	27.4	22.3	1091	21	AAC40517 Arabidopsis thalia
13	27.4	22.3	1689	22	AAF59609 Human cell cycle a
14	27.2	22.1	9965	24	ABL33526 Human immune syste
15	27	22.0	363	24	ABI99532 Mouse ischaemic co
16	26.8	21.8	6827	19	AAV52193 Streptococcus pneu
17	26.6	21.6	1030	24	AAC34296 Arabidopsis thalia
18	26.6	21.6	81001	22	AAF30035 Human apolipoprote
19	26.4	21.5	1035	21	AAH51499 Human growth facto
20	26.4	21.5	534720	19	AAV30458 Rhizobium species
21	26.4	21.5	536165	19	AAV30459 Rhizobium species
22	26.2	21.3	2091	18	AAV25107 H. pylori cell env
23	26.2	21.3	2091	20	AAV5826 H. pylori outer me
24	26.2	21.3	2091	24	AAI70670 Helicobacter pylor
25	26.2	21.3	2091	24	AAI70671 Helicobacter pylor
26	26.2	21.3	2114	23	ABL10737 Drosophila melanog
27	26.2	21.3	5581	23	ABL10736 Drosophila melanog
28	26.2	21.3	9012	23	ABL06304 Drosophila melanog
29	26.2	21.3	16428	22	AAK81010 Human immune/haema
30	26.2	21.3	31834	22	AAK81009 Human immune/haema
31	26.2	21.3	32035	22	AAK81008 Human immune/haema
32	26	21.1	395	22	AAI83339 Human polynucleoti
33	26	21.1	1006	21	AAH44726 Human secreted exp
34	26	21.1	2224	22	AAH06354 Human vasa cDNA.
35	25.8	21.0	2344	19	AAV20655 Open reading frame
36	25.8	21.0	3644	23	ABL09923 Drosophila melanog
37	25.8	21.0	6085	23	ABL09922 Drosophila melanog
38	25.6	20.8	300	21	AAH01389 Human colon cancer
39	25.6	20.8	421	22	AAH1648 Human polynucleoti
40	25.6	20.8	583	22	AAH11200 Human brain expres
41	25.6	20.8	2975	23	ABL10608 Drosophila melanog
42	25.6	20.8	3032	13	AAQ25325 Trichoderma reesei
43	25.6	20.8	3033	21	AAAG3953 DNA encoding a bet
44	25.6	20.8	3033	21	AAZ38879 Trichoderma reesei
45	25.6	20.8	4321	23	ABL18479 Drosophila melanog

ALIGNMENTS

RESULT 1

AAH22395

ID : AAH22395 standard; DNA; 123 BP.

AC AAH22395;

XX

XX

DT 22-AUG-2001 (first entry)

XX

DE Human rac1 genomic fragment probable 3' UTR SEQ ID NO:1.

XX

KW Identification; toxic; hepatotoxic; differential gene expression;

KW NSAID; non-steroidal antiinflammatory drug; ds.

XX

OS Homo sapiens.

XX

PN WO200138579-A2.

XX

PD 31-MAY-2001.

XX

PF 21-NOV-2000; 2000WO-US32049.

XX

PR 22-NOV-1999; 99US-0166923.

PR 18-FEB-2000; 2000US-0183531.

PR 20-NOV-2000; 2000US-0717321.

XX (CURA-) CURAGEN CORP.

XX

PI Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;

XX

XX WPI; 2001-355948/37.

DR

XX Screening hepatotoxic agent comprises contacting test cell population

PT

PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels
 XX
 PS Disclosure; Page 7; 76pp; English.

XX The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test cell
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to
 CC the expression of the nucleic acid sequence in a reference cell
 CC population comprising at least one cell whose exposure status to a
 CC hepatotoxic agent is known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.

XX Sequence 123 BP; 36 A; 28 C; 30 G; 29 T; 0 other;

Query Match 100.0%; Score 123; DB 22; Length 123;
 Best Local Similarity 100.0%; Pred. No. 8.6e-36;
 Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAGTTTCTAGTGTGCGAAGGCCCAACACACGTGTTCTTGCAGTGAG 60
 DB 1 CAATTGAAAAGTTTCTAGTGTGCGAAGGCCCAACACACGTGTTCTTGCAGTGAG 60
 QY 61 TTAGGTGTACAGACGGCTTAGCCTAGCGCTTGACAGACCTCACAGACCCAAAGGT 120
 DB 61 TTAGGTGTACAGACGGCTTAGCCTAGCGCTTGACAGACCTCACAGACCCAAAGGT 120
 QY 121 ACC 123
 DB 121 ACC-123

RESULT 2
 AAH22399/c
 ID AAH22399 standard; DNA; 348 BP.
 AC AAH22399;
 XX
 DT 22-AUG-2001 (first entry)
 XX Human rac1 gene related nucleotide sequence #3.
 XX
 KW Identification; toxic; hepatotoxic; differential gene expression;
 KW NSAID; non-steroidal antiinflammatory drug; ds.
 XX
 OS Homo sapiens.
 XX WO200138579-A2.
 XX
 PD 31-MAY-2001.
 XX
 PF 21-NOV-2000; 2000WO-US32049.
 XX
 PR 22-NOV-1999; 99US-0166923.
 PR 18-FEB-2000; 2000PS-0183531.
 PR 20-NOV-2000; 2000US-0717321.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
 XX WPI; 2001-355948/37.
 XX
 PT Screening hepatotoxic agent comprises contacting test cell population

PT expressing RISKMARKER or INJURYMARKER with agent, comparing expression
 PT with reference population and identifying difference in expression
 PT levels
 XX
 PS Disclosure; Page 8-9; 76pp; English.

XX The present invention describes a method of screening a test agent for
 CC hepatotoxicity. The method comprises: (a) providing a test cell
 CC population comprising a cell capable of expressing one or more nucleic
 CC acid sequences selected from the group consisting of RISKMARKER 1-8
 CC and INJURYMARKER 1-10; (b) contacting the test cell population with a
 CC test agent; (c) measuring expression of one or more of the nucleic
 CC acid sequences in the test cell population; (d) comparing the
 CC expression of the nucleic acid sequence in the test cell population to
 CC the expression of the nucleic acid sequence in a reference cell
 CC population comprising at least one cell whose exposure status to a
 CC hepatotoxic agent is known; and (e) identifying a difference in
 CC expression levels of the RISKMARKER or INJURYMARKER sequences, if
 CC present, in the test cell population and reference cell population.
 CC The method is useful for identifying a hepatotoxic agent. The present
 CC sequence is given in the exemplification of the present invention.

XX Sequence 348 BP; 103 A; 76 C; 67 G; 102 T; 0 other;

Query Match 100.0%; Score 123; DB 22; Length 348;
 Best Local Similarity 100.0%; Pred. No. 1.3e-35;
 Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAAGTTTCTAGTGTGCGAAGGCCCAACACACGTGTTCTTGCAGTGAG 60
 DB 212 CAATTGAAAAGTTTCTAGTGTGCGAAGGCCCAACACACGTGTTCTTGCAGTGAG 153
 QY 61 TTAGGTGTACAGACGGCTTAGCCTAGCGCTTGACAGACCTCACAGACCCAAAGGT 120
 DB 152 TTAGGTGTACAGACGGCTTAGCCTAGCGCTTGACAGACCTCACAGACCCAAAGGT 93
 QY 121 ACC 123
 DB 92 ACC 90

RESULT 3
 AAH22396
 ID AAH22396 standard; DNA; 1266 BP.
 AC AAH22396;
 XX
 DT 22-AUG-2001 (first entry)
 XX Human rac1 contig SEQ ID NO:2.
 XX
 KW Identification; toxic; hepatotoxic; differential gene expression;
 KW NSAID; non-steroidal antiinflammatory drug; ds.
 XX
 OS Homo sapiens.
 XX WO200138579-A2.
 XX
 PD 31-MAY-2001.
 XX
 PF 21-NOV-2000; 2000WO-US32049.
 XX
 PR 22-NOV-1999; 99US-0166923.
 PR 18-FEB-2000; 2000PS-0183531.
 PR 20-NOV-2000; 2000US-0717321.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 XX Gould-Rothberg BE, Dipippo VA, Ramseh TM, Gerwein RW;
 XX WPI; 2001-355948/37.
 XX
 PT Screening hepatotoxic agent comprises contacting test cell population

expressing RISKMARKER or INJURYMARKER with agent, comparing expression levels with reference population and identifying difference in expression

Disclosure; Page 7; 76pp; English.

The present invention describes a method of screening a test agent for hepatotoxicity. The method comprises: (a) providing a test cell population comprising a cell capable of expressing one or more nucleic acid sequences selected from the group consisting of RISKMARKER 1-8 and INJURYMARKER 1-10; (b) contacting the test cell population with a test agent; (c) measuring expression of one or more of the nucleic acid sequences in the test cell population; (d) comparing the expression of the nucleic acid sequence in the test cell population to the expression of the nucleic acid sequence in an reference cell population comprising at least one cell whose exposure status to a hepatotoxic agent is known; and (e) identifying a difference in expression levels of the RISKMARKER or INJURYMARKER sequences, if present, in the test cell population and reference cell population. The method is useful for identifying a hepatotoxic agent. The present invention is given in the exemplification of the present invention.

SQ Sequence 1266 BP; 385 A; 258 C; 285 G; 338 T; 0 other;

Query Match 100.0%; Score 123; DB 22; Length 1266;

Best Local Similarity 100.0%; Pred. No. 2.1e-35;

Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAATTGAAAGTGTCTAGTGTGCGAAGGCCCAACACATGCTTGTGCGAGTGAG 60

DB 146 CAATTGAAAGTGTCTAGTGTGCGAAGGCCCAACACATGCTTGTGCGAGTGAG 205

QY 61 TTAGTGTGTACAGACGGCGTTAGCTAGCGCTTGACAGAACTCAGACACCCAAAGGT 120

DB 206 TTAGTGTGTACAGACGGCGTTAGCTAGCGCTTGACAGAACTCAGACACCCAAAGGT 265

QY 121 ACC 123

DB 266 ACC 268

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PR 09-FEB-1999; 99US-0119286.

PR 11-FEB-1999; 99US-0119998.

PR 11-FEB-1999; 99US-0119759.

XX (RIGE-) RIGEL PHARM INC.

XX Luo Y;

PI WPI; 2000-482908/42.

XX New nucleic acids encoding Exo proteins which are useful in the

XX diagnosis, treatment or prevention of exocytosis-mediated disorders

XX such as asthma, inflammation and allergies

XX Disclosure; Page 271-272; 305pp; English.

XX The present sequence encodes a polypeptide which is associated with

XX the exocytosis pathway. CDNA molecules encoding proteins involved in

XX exocytosis have been isolated by yeast one-hybrid and two-hybrid

XX screening. Novel proteins, termed Exo proteins, have been identified that

XX interact with known exocytosis-associated proteins such as GS27, alpha

XX snap, unc18-1, vamp3, snap-23, and the rab family of proteins.

XX Exo proteins and their agonists and antagonists are useful in the

XX diagnosis, treatment or prevention of exocytosis-mediated disorders

XX such as asthma, inflammation, allergies, Chediak-Higashi Syndrome

XX (CHS), Alzheimer's disease, Parkinson's disease, Huntington's disease,

XX diabetes, digestion disorders and wound healing disorders.

XX The nucleic acids, antagonists or agonists of Exo proteins are useful

XX in gene therapy. The nucleic acids are also useful for generating

XX transgenic or knock-out animals which can be used in the

XX development and screening of therapeutically useful reagents.

XX SQ Sequence 422 BP; 95 A; 92 C; 83 G; 148 T; 4 other;

XX Query Match 86.7%; Score 106.6; DB 21; Length 422;

XX Best Local Similarity 91.8%; Pred. No. 1.6e-29;

XX Matches 112; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CAATTGAAAGTGTCTAGTGTGCGAAGGCCCAACACATGCTTGTGCGAGTGAG 60

DB 378 CAATTGAAAGTGTCTAGTGTGCGAAGGCCCAACACATGCTTGTGCGAGTGAG 319

QY 61 TTAGTGTGTACAGACGGCGTTAGCTAGCGCTTGACAGAACTCAGACACCCAAAGGT 120

DB 318 TTAGTGTGTACAGACGGCGTTAGCTAGCGCTTGACAGAACTCAGACACCCAAAGGT 259

QY 121 AC 122

DB 258 AC 257

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX DR WPI: 2001-639362/73.
 XX DR P-PSDB; AEC01382.
 XX DT New isolated polynucleotide and encoded polypeptides, useful in
 XX DT diagnostics, forensics, gene mapping, identification of mutations
 XX DT responsible for genetic disorders or other traits and to assess
 XX DT biodiversity -
 XX PI
 XX PS Claim 1; SEQ ID No 1373; 103pp; English.
 XX CC The invention relates to isolated polynucleotide (I) and
 XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 XX CC and gene mapping, and in recombinant production of (II). The
 XX CC polynucleotides are also used in diagnostics as expressed sequence tags
 XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
 XX CC to restore normal activity of (II) or to treat disease states involving
 XX CC (II). (II) is useful for generating antibodies against it, detecting or
 XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
 XX CC a food supplement. (II) and its binding partners are useful in medical
 XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
 XX CC disorders involving aberrant protein expression or biological activity.
 XX CC The polypeptide and polynucleotide sequences have applications in
 XX CC diagnostics, forensics, gene mapping, identification of mutations
 XX CC responsible for genetic disorders or other traits to assess biodiversity
 XX CC and to produce other types of data and products dependent on DNA and
 XX CC amino acid sequences. AAS64197-AAS94584 represent novel human
 XX CC diagnostic coding sequences of the invention.
 XX CC Note: The sequence data for this patent did not appear in the printed
 XX CC specification, but was obtained in electronic format directly from WIPO
 XX CC at ftp.wipo.int/pub/published_pct_sequences.
 XX DT
 XX DT Sequence 1318 BP; 346 A; 268 C; 279 G; 424 T; 1 other;
 Query Match 64.7%; Score 79.6; DB 23; Length 1318;
 Best Local Similarity 83.1%; Pred. No. 2.4e-19;
 Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
 QY 1 CAATTGAAAAGTTGTTCTAGTGGTGAAGAGGCCCAACACTGTGTTCTGCCAGTGAG 60
 DB 1221 CAATTGAAAAGTTGTTCTAGTGGTGAAGAGGCCCAACACTGTGTTCTGCCAGTGAG 1162
 QY 61 TTAGGTGTACAGAACGCGCTTACGACTAGCGC--TTGACAGAACCTTCACAGACCCCAAG 118
 DB 1161 TTAGGTGTACAGAACATCGTCAGCACTAGCAGATTACAGAACTTCACAGACCCCAAG 1102
 QY 119 GTAC 122
 DB 1101 GAAC 1098
 RESULT 8
 AAF18236/c
 ID AAF18236 standard; DNA; 2051 BP.
 XX AC AAF18236;
 XX DT

14-MAR-2001 (first entry)
 Lung cancer associated polynucleotide sequence SEQ ID 255.
 Human; lung cancer associated protein; neuroprotective; cytostatic;
 cardioactive; immunomodulatory; muscular active; vulnerary;
 gastrointestinal; nephrotropic; antineoplastic; gynecological;
 proliferative disorder; wound healing; infectious disease; ds.
 Homo sapiens.
 WO200055180-A2.
 21-SEP-2000.
 08-MAR-2000; 2000WO-US05918.
 12-MAR-1999; 99US-0124270.
 (HUMA-) HUMAN GENOME SCI INC.
 (ROSE/) ROSEN C A.
 Ruben SM;
 WPI: 2000-587514/55.
 P-PSDB; AAB58360.
 Lung cancer associated gene sequences, referred to as lung cancer
 antigens, useful for treatment, prevention, and diagnosis of disorders
 such as lung cancer -
 Claim 1; Page 716-717; 1425pp; English.
 Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 associated proteins represented in AAB58106 - AAB58548. Lung cancer
 associated proteins and polynucleotide sequences, their agonists, and
 antagonists may have neuroprotective; cytostatic; cardioactive;
 immunomodulatory; muscular active general; vulnerary; gastrointestinal
 general; nephrotropic; antineoplastic; gynecological; or antibacterial
 activity. The invention also includes antibodies specific for the
 protein or polynucleotide sequences. The lung cancer associated
 polynucleotide sequences may be used for detection of lung cancer,
 chromosome identification, as chromosome markers, and for numerous other
 diagnostic or research purposes. The proteins may be used to treat
 disorders such as neural, immune, muscular, reproductive,
 gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 disorders. The proteins may also be used in the treatment of wounds and
 infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
 peptide AAB58549 are used in the course of the invention for the
 identification and characterisation of the polynucleotide and protein
 sequences.
 Sequence 2051 BP; 570 A; 430 C; 433 G; 612 T; 6 other;
 Query Match 64.7%; Score 79.6; DB 21; Length 2051;
 Best Local Similarity 83.1%; Pred. No. 2.9e-19;
 Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
 QY 1 CAATTGAAAAGTTGTTCTAGTGGTGAAGAGGCCCAACACTGTGTTCTGCCAGTGAG 60
 DB 1896 CAATTGAAAAGTTGTTCTAGTGGTGAAGAGGCCCAACACTGTGTTCTGCCAGTGAG 1837
 QY 61 TTAGGTGTACAGAACGCGCTTACGACTAGCGC--TTGACAGAACCTTCACAGACCCCAAG 118
 DB 1836 TTAGGTGTACAGAACATCGTCAGCACTAGCAGATTACAGAACCTTCACAGACCCCAAG 1777
 QY 119 GTAC 122
 DB 1776 GAAC 1773
 RESULT 9

AA07296/c
ID AAL07296 standard; DNA; 3740 BP.
XX AC AAL07296;
XX DT 21-NOV-2001 (first entry)
XX Human reproductive system related antigen DNA SEQ ID NO: 9984.
DE Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX Homo sapiens.
XX WO200155320-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01339.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231124.
PR 08-SEP-2000; 2000US-0231124.
PR 08-SEP-2000; 2000US-0231143.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234984.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/alcalteric bowel
CC diseases. The present sequence is a gene of the invention.

Query Match	22.1%	Score 27.2;	DB 24;	Length 9965;
Best Local Similarity	58.8%	Pred. No. 13;		
Matches 47;	Conservative	0;	Mismatches 33;	Indels 0;

QY

4 TTGAAAAAGTTTGTCTAGTGTCGAAAGGCCAACACATGTGTTCTTGCACGTGGATT 63
|| ||||||| ||| | ||||||| ||| | ||||||| ||| | ||||||| ||| | |||||||

D8

3470 TTTTTAAAAAATTTTTTATAGTGGTAGAAGCGAGGAGATAATCTTTGTAGTTTTTGGCCTTA 3529
|| ||||||| ||| | ||||||| ||| | ||||||| ||| | ||||||| ||| | |||||||

Qy 64 GGTGTGTACAGAAACGGCGTTA 83
||| ||| | | | |
Db 3530 GGTGATAGTTAGGGAAGGTA 3549

RESULT 15
ABI99532

ABI99532
ID ABI99532 standard; cDNA: 363 BP.

XX
AC
ABT99532.

DT 07-MAR-2002 (first entry)

Mouse ischaemic condition related cDNA sequence SEO ID NO:530.

AA
KW
AA
KW
KW
KW

OS Mus musculus.

PN WO200188188-A2.

22-NOV-2001.

XX
PF
18-MAY-2001 2001WO-JP04192

19-MAY-2000 2000TD-0145077

XX

PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

DR WPI; 2002-034733/04.

PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -

PS Claim 2: page 1454: 2690pp: English.

The present invention describes a method for examining ischaemic conditions, comprising measuring the expression levels of particular genes (1) in a test sample or determining the expression profile of a gene group in the sample comprising genes selected from (1). The method is useful for examining the ischaemic condition (e.g. compressive ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring expression levels of particular genes (AB199202 to AB199912, encoding the protein sequences in AB57020 to AB57374) or by determining the expression profile of a gene group comprising these genes. The expression levels or expression profiles produced by these genes are used as an indicator when screening for ischaemic condition-improving drugs or therapeutics for ischaemic diseases. AB199913 and AB199914

CC represent PCR primers for a mouse ischaemic condition related sequence,
 CC which are used in the exemplification of the present invention.

XX
 SQ Sequence 363 BP; 115 A; 57 C; 93 G; 98 T; 0 other;

Query Match 22.0%; Score 27; DB 24; Length 363;

Best Local Similarity 66.1%; Pred. NO. 4.4;

Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 8 AAAAAAGTTTCTTCTAGTGGTCGAAGGCCCAACACTGTGTCTTCCACGTGACTTAGGT 66

||| ||||| ||||| || || || ||||| ||||| ||||| ||||| ||

Db 187 AAAGGGTTTGTGCAAGTGGATGATGCGAGAAAAATTGTGTCTGCTGGGTGTCTGT 245

Search completed: October 26, 2002, 04:00:35

Job time : 183 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 26, 2002, 03:01:19 ; Search time 1846 Seconds
(without alignments)
1394.347 Million cell updates/sec

Title: US-09-717-321a-1

Perfect score: 123

Sequence: 1 caattgaaaaagtgttgc.....ctcagagaccacaaagtacc 123

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl.:

1: gb_ba.:

2: gb_htg.:

3: gb_in.:

4: gb_om.:

5: gb_ov.:

6: gb_pat.:

7: gb_ph.:

8: gb_pl.:

9: gb_pr.:

10: gb_ro.:

11: gb_sts.:

12: gb_sy.:

13: gb_un.:

14: gb_vl.:

15: em_ba.:

16: em_fun.:

17: em_hum.:

18: em_in.:

19: em_mu.:

20: em_om.:

21: em_or.:

22: em_ov.:

23: em_pat.:

24: em_ph.:

25: em_pl.:

26: em_ro.:

27: em_sts.:

28: em_un.:

29: em_vl.:

30: em_htg_hum.:

31: em_htg_inv.:

32: em_htg_other.:

33: em_htgo_inv.:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

1	123	100.0	123	6	AX163737	Sequence
2	123	100.0	348	6	AX163753	Sequence
3	123	100.0	1266	6	AX163738	Sequence
4	107.6	87.5	2319	10	BC003828	Mus muscu
5	107.6	87.5	269081	2	AC068493	Mus muscu
6	79.6	64.7	352	6	AX163754	Sequence
7	79.6	64.7	455	11	G26995	human sys S
8	79.6	64.7	1232	9	HMP02ST9	D25274 Homo sapien
9	79.6	64.7	2315	9	AK054993	Homo sapi
10	79.6	64.7	28567	9	HSAL32695	Homo sapi
11	79.6	64.7	212827	9	AC009412	Homo sapi
12	77.6	63.1	262	11	G31709	SWSS22233 Er
13	74.8	60.8	137625	9	AC104663	Homo sapi
14	74.8	60.8	159960	30	AC024963	Homo sapi
15	48	39.0	101584	9	CNS01DS5	BAC seque
16	48	39.0	155943	9	AC012364	Homo sapi
17	38.2	31.1	110816	9	AC002404	Human Chr
18	38.2	31.1	186218	2	AC079169	Homo sapi
19	33.2	27.0	1485	5	AF029250	Danlo rer
20	32.4	26.3	110000	2	AC107387_0	Homo sapi
21	32.4	26.3	152747	9	AC092576	Homo sapi
22	32.4	26.3	322593	2	AC107084	Homo sapi
23	31.4	25.5	605	6	AX341407	Sequence
24	31.2	25.4	90083	2	AC103167	Rattus no
25	30.8	25.0	97955	2	AC108543	Rattus no
26	30.6	24.9	166081	2	AL626786	Mus muscu
27	30.6	24.9	167100	9	AC008875	Homo sapi
28	30.6	24.9	236029	2	AC019232	Homo sapi
29	30.4	24.7	99098	2	AC095793	Rattus no
30	30.4	24.7	147098	2	AC097183	Rattus no
31	30.2	24.6	253	11	A0026917	Rattus no
32	30	24.4	1286	8	EGU37794	Eucalyptus
33	30	24.4	86945	2	AC095700	Rattus no
34	30	24.4	91940	9	AL445674	Human DNA
35	30	24.4	183378	2	AL607151	Mus muscu
36	30	24.4	224712	2	AL606969	Mus muscu
37	29.8	24.2	307	11	HSB016XG9	H.sapiens
38	29.6	24.1	68108	2	AC090474	Homo sapi
39	29.4	23.9	509	11	G63194	SHGC-140826
40	29.4	23.9	82024	2	AC023210	Homo sapi
41	29.4	23.9	143802	9	AL161431	Human DNA
42	29.4	23.9	175504	9	AC019196	Homo sapi
43	29.4	23.9	273403	9	AC011498	Homo sapi
44	29.2	23.7	124645	9	HUAC004638	Homo sapi
45	29.2	23.7	124998	9	AC026397	Homo sapi

ALIGNMENTS

RESULT 1	AX163737	Sequence 1 from Patent WO0138579.	123_bp	DNA	linear	PAT 22-JUN-2001
LOCUS	AX163737	Sequence 1 from Patent WO0138579.				
DEFINITION	AX163737					
ACCESSION	AX163737.1	GI:14544843				
VERSION						
KEYWORDS						
SOURCE	Norway rat.					
ORGANISM	Rattus norvegicus					
REFERENCE						
AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
TITLE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;					
JOURNAL	Rattus.					
FEATURES	1 (bases 1 to 123)					
source	Gould-Rothberg, B.E., Dipippo, V.A., Ramsch, T.M. and Gerwein, R.W.					
	Method of identifying toxic agents using nsaid-induced differential					
	gene expression in liver					
	Patent: WO 0138579-A 1 31-MAY-2001;					
	Curagen Corporation (US)					
	Location/Qualifiers					
	1..123					
	/organism="Rattus norvegicus"					
	/db_xref="taxon:10116"					
BASE COUNT	36 a	28 c	30 g	29 t		

ORIGIN
Query Match 100.0%; Score 123; DB 6; Length 123;
Best Local Similarity 100.0%; Pred. No. 6.9e-33;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 60
|||||
Db 1 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 60
|||||

Qy 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGTTGACAGAACCTTCAGACCCCAAGGT 120
|||||
Db 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGTTGACAGAACCTTCAGACCCCAAGGT 120
|||||

Qy 121 ACC 123
|||
Db 121 ACC 123

RESULT 2
AX163753/c
LOCUS AX163753 348 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 17 from Patent WO0138579.
ACCESSION AX163753
VERSION AX163753.1 GI:14544859
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 348)
AUTHORS Gould-Rothberg, B.E., Dippio, V.A., Ramse, T.M. and Gerwein, R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential
gene expression in liver
JOURNAL Patent: WO 0138579-A 17 31-MAY-2001;
Curagen Corporation (US)
FEATURES
source 1. .348
/db_xref="taxon:10116"
BASE COUNT 103 a 76 c 67 g 102 t
ORIGIN

Query Match 100.0%; Score 123; DB 6; Length 348;
Best Local Similarity 100.0%; Pred. No. 7.6e-33;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 60
|||||
Db 212 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 153
|||||

Qy 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGTTGACAGAACCTTCAGACCCCAAGGT 120
|||||
Db 152 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGTTGACAGAACCTTCAGACCCCAAGGT 93
|||||

Qy 121 ACC 123
|||
Db 92 ACC 90

RESULT 3
AX163738
LOCUS AX163738 1266 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 2 from Patent WO0138579.
ACCESSION AX163738
VERSION AX163738.1 GI:14544844
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 1266)
AUTHORS Gould-Rothberg, B.E., Dippio, V.A., Ramse, T.M. and Gerwein, R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential
gene expression in liver
JOURNAL Patent: WO 0138579-A 2 31-MAY-2001;
Curagen Corporation (US)
FEATURES
source 1. .1266
/db_xref="taxon:10116"
BASE COUNT 385 a 258 c 285 g 338 t
ORIGIN

Query Match 100.0%; Score 123; DB 6; Length 1266;
Best Local Similarity 100.0%; Pred. No. 8.6e-33;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 60
|||||
Db 146 CAATTGAAAAGTTTGTCTAGTGGTGAAGGCCCAACACTGTGTCTTCCAGTGAG 205
|||||

Qy 61 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGTTGACAGAACCTTCAGACCCCAAGGT 120
|||||
Db 206 TTAGGTTGTACAGAACGGCGTTAGCACTAGCGTTGACAGAACCTTCAGACCCCAAGGT 265
|||||

Qy 121 ACC 123
|||
Db 266 ACC 268

RESULT 4
BC003828/c
LOCUS BC003828 2319 bp mRNA linear ROD 12-JUL-2001
DEFINITION Mus musculus, ras-related C3 botulinum toxin substrate 1 (rho
family, small GTP binding protein rac1), clone MGC:6235
IMAGE:3593957, mRNA, complete cds.
ACCESSION BC003828
VERSION BC003828.1 GI:13277917
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2319)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NCI-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcdpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 11 Row: e Column: 22
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 53885.

FEATURES
source Location/Qualifiers
1. .2319
/db_xref="taxon:10090"

```
/clones="MGC:6235 IMAGE:3593957"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
198. .776
/codon_start=1
/product="ras-related C3 botulinum toxin substrate 1 (rho
family, small GTP binding protein Rac1)"
/protein_id="AA03828.1"
/db_xref="GI:13277918"
/translacion="MQAIKCVVGDGAVGKTCLLISVTTNAPGEYIPTVDNYSANV
HVDGKPNGLMDTAGEDYDLRPLSPQDTVELICFSLVSPASFNRAKRWPEYR
HFCNPITLIVGTGTLIDLDKDDTKLEKRLKLPITYPQGLAMAKEIGAVKYLECSAL
TORGLKRVDEAIRAVLCPPVKKRKKLL"
BASE COUNT 603 a 573 c 503 g 640 t
ORIGIN
Query Match 87.5%; Score 107.6; DB 10; Length 2319;
Best Local Similarity 92.6%; Pred. No. 2.7e-27;
Matches 113; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 CAATTGAAAAGTTGTTCTAGTGGTGAAGAGCCCAACTGTGTTCTGCCAGTGAG 60
Db 2173 CAATTGAAAAGTTGTTCTAGTGGTGAAGAGCCCAACTGTGTTCTGCCAGTGAG 2114
QY 61 TTAGTGTGTACAGAGCGGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCCAAGGT 120
Db 2113 TTAGTGTGTACAGAGCGGTTAGCACTAGCGCTTGACAGAACCTCACAGACCCCAAGGA 2054
QY 121 AC 122
Db 2053 AC 2052
RESULT 5
AC068493/c
LOCUS
DEFINITION Mus musculus clone RP23-76K1 strain C57BL6/J, WORKING DRAFT
SEQUENCE, 26 unordered pieces.
ACCESSION AC068493
VERSION AC068493.10 GI:15148081
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 269081)
Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 269081)
Montgomery,K.T., Grills,G., Han,J., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
Direct Submission
Submitted (03-MAY-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Aug 11, 2001 this sequence version replaced gi:14993654.
-----Genome Center
Center: Harvard Partners Genome Center
Center Code: HPGC
Web site:
http://wchanning.bwh.harvard.edu:9088/hpcpg/jsp/hpcpg/Sequence/mous
e.html
Contact: gntm@capecod.bwh.harvard.edu
-----Summary Statistics
Center project name: ABN
Sequencing vector: pUC18; L08752
```

```
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 256302 at least Q20
*Consensus quality: 254124 at least Q30
*Consensus quality: 250276 at least Q40
Estimated insert size: agarose-FP - N/A
**Estimated insert size: 268581 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality/coverage: 6.2 x in Q20 bases; sum-of-contigs estimation
-----
NOTE: This is a 'working draft' sequence. It currently
consists of 26 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 51730: contig of 51730 bp in length
* 51731: gap of unknown length
* 51751: contig of 30673 bp in length
* 82423: gap of unknown length
* 82424: gap of unknown length
* 82444: contig of 35715 bp in length
* 118158: contig of 35715 bp in length
* 118159: gap of unknown length
* 118178: gap of unknown length
* 118179: contig of 21263 bp in length
* 139441: gap of unknown length
* 139461: gap of unknown length
* 139462: contig of 21177 bp in length
* 160638: contig of 21177 bp in length
* 160639: gap of unknown length
* 160658: gap of unknown length
* 160659: contig of 16953 bp in length
* 177611: contig of 16953 bp in length
* 177612: gap of unknown length
* 190991: contig of 13360 bp in length
* 190992: gap of unknown length
* 191011: gap of unknown length
* 191012: contig of 9117 bp in length
* 200128: contig of 9117 bp in length
* 200148: gap of unknown length
* 200149: contig of 11734 bp in length
* 211882: contig of 11734 bp in length
* 211902: gap of unknown length
* 211903: contig of 8871 bp in length
* 220774: gap of unknown length
* 220794: contig of 8850 bp in length
* 229643: contig of 8850 bp in length
* 229644: gap of unknown length
* 229664: gap of unknown length
* 238339: contig of 8676 bp in length
* 238340: gap of unknown length
* 238359: gap of unknown length
* 238360: contig of 6489 bp in length
* 244849: gap of unknown length
* 244868: gap of unknown length
* 249725: contig of 4857 bp in length
* 249726: gap of unknown length
* 249745: gap of unknown length
* 252523: contig of 2778 bp in length
* 252524: gap of unknown length
* 252543: gap of unknown length
* 255202: contig of 4977 bp in length
* 257540: gap of unknown length
* 260396: contig of 2856 bp in length
* 260397: gap of unknown length
* 260416: gap of unknown length
* 262709: contig of 2293 bp in length
* 262710: gap of unknown length
* 262729: gap of unknown length
* 263934: contig of 1205 bp in length
* 263935: gap of unknown length
* 263955: gap of unknown length
* 264196: contig of 242 bp in length
* 264217: gap of unknown length
* 264219: contig of 413 bp in length
* 264629: gap of unknown length
* 264630: gap of unknown length
* 264649: gap of unknown length
* 265608: contig of 959 bp in length
* 265628: gap of unknown length
* 265629: contig of 1307 bp in length
* 266935: gap of unknown length
* 266936: gap of unknown length
* 266955: gap of unknown length
* 268335: contig of 1380 bp in length
* 268336: gap of unknown length
* 268356: gap of unknown length
* 268357: contig of 194 bp in length
* 268550: gap of unknown length
* 268551: gap of unknown length
* 269081: contig of 512 bp in length.
Location/Qualifiers
1. 269081
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
```

/clone="RP23-76K1"
/sex="male"
1. .51730
/note="assembly_name:Contig138"
51751. .82423
/note="assembly_name:Contig137"
82444. .118158
/note="assembly_name:Contig136"
118179. .139441
/note="assembly_name:Contig135"
139462. .160638
/note="assembly_name:Contig134"
160659. .177611
/note="assembly_name:Contig133"
177632. .190991
/note="assembly_name:Contig132"
191012. .200128
/note="assembly_name:Contig131"
200149. .211882
/note="assembly_name:Contig130"
211903. .220773
/note="assembly_name:Contig129"
220794. .229643
/note="assembly_name:Contig128"
229664. .238339
/note="assembly_name:Contig127"
238360. .244848
/note="assembly_name:Contig126"
244869. .249725
/note="assembly_name:Contig125"
249746. .252523
/note="assembly_name:Contig124"
vector_side:left
252544. .257520
/note="assembly_name:Contig123"
257541. .260396
/note="assembly_name:Contig122"
260417. .262709
/note="assembly_name:Contig121"
262730. .263934
/note="assembly_name:Contig120"
263955. .264196
/note="assembly_name:Contig119"
264217. .264629
/note="assembly_name:Contig118"
264650. .265608
/note="assembly_name:Contig117"
265629. .266935
/note="assembly_name:Contig116"
266956. .268335
/note="assembly_name:Contig115"
268356. .268549
/note="assembly_name:Contig114"
268570. .269081
/note="assembly_name:Contig113"
BASE COUNT 69726 a 66027 c 64672 g 67724 t 932 others
ORIGIN

Query Match 87.5%; Score 107.6; DB 2; Length 269081;
Best Local Similarity 92.6%; Pred. No. 4.2e-27;
Matches 113; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 1 CAATTGAAAGAACTTTGTTCTAGTGTGCGAAGCCCAACACAGTGTCTTCTGCGAGTGAG 60
|||||
Db 63360 CAATTGAAAGAACTTTGTTCTAGTGTGCGAAGCCCAACACAGTGTCTTCTGCGAGTGAG 63301
QY 61 TTAGGTTGTACAGACGGCTTACGACTAGCGCTTGACAGAACTTCACAGACCCCAAGGT 120
|||||
Db 63300 TTAGGTTGTACAGACGGCTTACGACTAGCGCTTGACAGAACTTCACAGACCCCAAGGT 63241
QY 121 AC 122
||

Db 63240 AC 63239
RESULT 6
LOCUS AX163754/c
DEFINITION Sequence 18 from Patent WO0138579.
ACCESSION AX163754
VERSION AX163754.1 GI:14544860
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 352)
AUTHORS Gould-Rothberg,B.E., DiPippo,V.A., Rameh,T.M. and Gerwein,R.W.
TITLE Method of identifying toxic agents using nsaid-induced differential
JOURNAL gene expression in liver
Patent: WO 0138579-A 18 31-MAY-2001;
Curagen Corporation (US)
FEATURES
source Location/Qualifiers
1. .352
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 86 a 77 c 72 g 117 t
ORIGIN
Query Match 64.7%; Score 79.6; DB 6; Length 352;
Best Local Similarity 83.1%; Pred. No. 2e-17;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;
QY 1 CAATTGAAAGAACTTTGTTCTAGTGTGCGAAGCCCAACACAGTGTCTTCTGCGAGTGAG 60
|||||
Db 216 CAATTGAAAGAACTTTGTTCTAGTGTGCGTGAAGGCTCCACGCTGATATCTGCGAGTGAG 157
QY 61 TTAGGTTGTACAGACGGCTTACGACTAGCGC--TTGACAGAACCTTCACAGACCCCAAG 118
|||||
Db 156 TTAAGTTGTACAGACATCGTCAGCAGTACGACAGTTTACAGAACCTTCACAGACCCCAAG 97
QY 119 GTAC 122
|||
Db 96 GAAC 93
RESULT 7
LOCUS G26995
DEFINITION human STS SHGC-31947, sequence tagged site.
ACCESSION G26995
VERSION G26995.1 GI:1375245
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 455)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL Myers,R.M.
COMMENT Unpublished
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu
Primer A: AAGTCATACATGCAAAACATATGCG
Primer B: GCGGAGATACAGCGTGG
STS size: 131
PCR Profile:
Initial incubation: 94 degrees C for 90 seconds

construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

source Location/Qualifiers
 1. .2315
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="BRACE2008958"
 /tissue_type="cerebellum"
 /clone_lib="BRACE2"
 /notes="cloning vector: pME18SFL3"
 BASE COUNT 603 a 531 c 501 g 680 t
 ORIGIN

Query Match 64.7%; Score 79.6; DB 9; Length 2315;

Best Local Similarity 83.1%; Pred. No. 2.4e-17;

Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

1 CAATTGAAAGTCTTCTAGTGGTGAAGGCCCAACACGTGTCTTCCAGTGAG 60
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 2219 CAATTGAAAGTCTTCTAGTGGTGAAGGCCCAACACGTGTCTTCCAGTGAG 2160
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 61 TTAGGTGTACAGACGGCTTAGCCTAGCGC--TTGACAGACCTCAGACCCCAAAG 118
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 DB 2159 TTAAGTGTACAGACATCTCAGCACTACACAGTTTACAGACCTCAGACCCCAAAG 2100
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 QY 119 GTAC 122
 ||||
 DB 2099 GAAC 2096

RESULT 10

HSAL32695/c 28567 bp DNA linear PRI 09-NOV-2000
 LOCUS Homo sapiens rai1 gene.
 DEFINITION Homo sapiens rai1 gene.
 ACCESSION AJ132695
 VERSION AJ132695.5 GI:8574037
 KEYWORDS alternative splicing; Alu; AluJo; AluSg1; AluSp; AluSx; Alu-rich; Cγ-rich; M1x; rai1 gene; rai1 protein; rai1 protein; repetitive sequence.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 28567)
 AUTHORS Matos, P., Skaug, J., Marques, B., Beck, S., Verissimo, F., Gespach, C., Boavida, M.G., Scherer, S.W. and Jordan, P.
 TITLE Small GTPase Rai1: Structure, Localization, and Expression of the Human Gene
 JOURNAL Biochem. Biophys. Res. Commun. 277 (3), 741-751 (2000)
 PUBMED 11062023
 REFERENCE 2 (bases 1 to 28567)
 AUTHORS Jordan, P.
 TITLE Direct Submission
 JOURNAL Submitted (03-FEB-1999) Jordan P., Centro de Genetica Humana, Laboratorio de Oncobiologia, Instituto Nacional de Saude 'Dr. Ricardo Jorge', Avenida Padre Cruz, 1649-016 Lisboa, PORTUGAL
 COMMENT On Jun 20, 2000 this sequence version replaced gi:7248282.
 Related sequence AI638561.

FEATURES

source Location/Qualifiers
 1. .28567
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7p22"
 /clone="H. NH0425A05, RPCI-11"
 repeat_region 1. .109
 /rpt_family="AluSg"
 repeat_region 125. .422
 /rpt_family="AluSc"
 repeat_region 441. .742

promoter /rpt_family="AluSq"
 743. .1226
 /gene="rai1"
 743. .28117
 /gene="rai1"
 911. .983
 /rpt_family="GC_rich"
 936
 /gene="rai1"
 /note="alternative transcription start site"
 1026
 /gene="rai1"
 /note="alternative transcription start site"
 1037
 /gene="rai1"
 /note="alternative transcription start site"
 1053
 /gene="rai1"
 /note="alternative transcription start site"
 1106. .1175
 /rpt_family="GC_rich"
 join(1227. .1261,13701. .13772,18413. .18530,22804. .22860,
 24268. .24330,26010. .26169,26458. .26588)
 /gene="rai1"
 /note="alternative"
 /codon_start=1
 /product="Rai1b protein"
 /protein_id="CAAI0733.6"
 /db_xref="GI:8574039"
 /db_xref="SPTREMBL:O95501"
 /translation="MOAIKCVVVDGAVGKTCLLISYTTNAPGCEIPTVFDNYSANV
 MVDGKPNVLGLWDTAGQEDYDLRLPLSYDPTGKDTSRGKDKPIADVFLICFS
 LVSPAFENVRHWYPEVHHCNPTFIILVGTGLDLRDDDTIEKLEKKLTPTIYPO
 GLAWAKEIGAVKYLECSALTQRLKTVFDEAIRAVLCPPPPVKKRKKRCLLL"
 <1227. .1261
 /gene="rai1"
 /number=1
 join(1227. .1261,13701. .13772,18413. .18530,24268. .24330,
 26010. .26169,26458. .26588)
 /gene="rai1"
 /codon_start=1
 /product="Rai1 protein"
 /protein_id="CAB53579.5"
 /db_xref="GI:8574038"
 /translation="MOAIKCVVVDGAVGKTCLLISYTTNAPGCEIPTVFDNYSANV
 MVDGKPNVLGLWDTAGQEDYDLRLPLSYDPTGKDTSRGKDKPIADVFLICFS
 LVSPAFENVRHWYPEVHHCNPTFIILVGTGLDLRDDDTIEKLEKKLTPTIYPO
 GLAWAKEIGAVKYLECSALTQRLKTVFDEAIRAVLCPPPPVKKRKKRCLLL"
 1262. .13700
 /gene="rai1"
 /number=1
 1390. .1419
 /rpt_family="GC_rich"
 1752. .1793
 /rpt_family="FLAM"
 1794. .2005
 /rpt_family="AluJo"
 2140. .2171
 /rpt_family="Simple_repeat"
 2666. .2972
 /rpt_family="AluSx"
 2980. .3267
 /rpt_family="AluSx"
 3288. .3289
 /rpt_family="(T)n"
 3802. .4102
 /rpt_family="AluJb"
 4122. .4386
 /rpt_family="AluSx"
 4538. .4719
 /rpt_family="LTR33A"
 4754. .5053
 /rpt_family="AluSq"

```

repeat_region 5297..5594
/rpt_family="AluSx"
repeat_region 5655..5757
/rpt_family="AluSx"
repeat_region 5758..6062
/rpt_family="AluSx"
repeat_region 6063..6198
/rpt_family="AluSx"
repeat_region 6222..6401
/rpt_family="AluSg/x"
repeat_region 7162..7238
/rpt_family="L1MC5"
repeat_region 7348..7367
/rpt_family="Simple_repeat"
repeat_region 7368..7656
/rpt_family="AluSx"
repeat_region 7880..8183
/rpt_family="AluSx"
repeat_region 8484..8778
/rpt_family="AluSx"
repeat_region 9071..9365
/rpt_family="AluSx"
repeat_region 10279..10583
/rpt_family="AluSp"
repeat_region 10609..10864
/rpt_family="L2"
repeat_region 10970..11276
/rpt_family="AluJo"
repeat_region 11276..11594
/rpt_family="AluJb"
repeat_region 11744..11825
/rpt_family="L2"
repeat_region 11764..11987
/rpt_family="MIR"
repeat_region 12989..13307
/rpt_family="AluSx"
exon 13701..13772
/gene="rac1"
Intron 13773..18412
/number=2
/gene="rac1"
repeat_region 13954..14235
/rpt_family="AluSx"
repeat_region 14807..14831
/rpt_family="Simple_repeat"
repeat_region 14836..15128
/rpt_family="AluSx"
repeat_region 15408..15715
/rpt_family="AluSx"
repeat_region 15872..15896
/rpt_family="AT-rich"
repeat_region 15902..16186
/rpt_family="AluSx"
repeat_region 16187..16314
/rpt_family="AluJo"
repeat_region 16413..16585
/rpt_family="Simple_repeat"
repeat_region 16586..16679
/rpt_family="FLAM_A"
repeat_region 16707..16810
/rpt_family="L1M4"
repeat_region 16811..17108
/rpt_family="AluSg"
repeat_region 17109..17345
/rpt_family="L1M4"
repeat_region 17377..17686
/rpt_family="AluSx"
repeat_region 17703..17729
/rpt_family="Simple_repeat"
exon 18413..18530
/gene="rac1"
/number=3

```

```

Intron 18531..22803
/gene="rac1"
/number=3
repeat_region 19287..19436
/rpt_family="MIR"
repeat_region 21161..21184
/rpt_family="Simple_repeat"
repeat_region 21185..21351
/rpt_family="AluSp"
misc_difference 21364..21365
/gene="rac1"
/number=2345 nt insertion sequence in RPI1 11-derived BAC
clones 425P05 and 18101"
repeat_region 21365..21403
/rpt_family="Alu"
repeat_region 21443..21720

Query Match 64.7% Score 79.6; DB 9; Length 28567;
Best Local Similarity 83.1%; Pred. No. 3e-17;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAGTTTCTTCTAGTGGTCGAAAGCCCAACACTGTGTCTTCCAGTGAG 60
Db 28012 CAATTTAAAAATTCCTCTAGTGGCTGAAGGCTCCACGCTGTTCTGCCAGTGAG 27953

QY 61 TTAGTTGTACAGAACATCGTCAGTACTAGCGGCTTACAGACCTCAGACCCCAAG 118
Db 27952 TTAAGTTGTACAGAACATCGTCAGTACTAGCGGCTTACAGACCTCAGACCCCAAG 27893

QY 119 GTAC 122
Db 27892 GAAC 27889

RESULT 11
AC009412/c 212827 bp DNA linear PRI 07-NOV-2001
LOCUS Homo sapiens BAC clone RP11-425P5 from 7, complete sequence.
DEFINITION AC009412
ACCESSION AC009412
VERSION AC009412.6 GI:14190769
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 212827)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 212827)
AUTHORS Hou, S., Maupin, R., Haakenson, W., Gregory, S., and Belter, E.
TITLE The sequence of Homo sapiens BAC clone RP11-425P5
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 212827)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 212827)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 212827)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 23, 2001 this sequence version replaced gi:13431187.
----- Genome Center

```

Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@wustl.edu
 ----- Summary Statistics -----
 Center project name: H_NH0425P05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
 VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTD-2195F2, 200 bp overlapp. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 212633 of RP11-425P5.

FEATURES

Source

1. .212827
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7"
 /clone="RP11-425P5"
 /clone_lib="RP11-11"

repeat_region

1623..1696

repeat_region

/rpt_family="MIR"

repeat_region

/rpt_family="Alu"

repeat_region

3006..3093

repeat_region

/rpt_family="L1"

repeat_region

3453..3552

repeat_region

/rpt_family="L2"

repeat_region

4924..5046

repeat_region

/rpt_family="ERV1"

repeat_region

5318..5693

repeat_region

/note="similar to EST BF996284 (NID:g12402607)"

repeat_region

5580..5845

repeat_region

/rpt_family="Alu"

repeat_region

5925..6277

repeat_region

/rpt_family="MIR"

repeat_region

6842..7041

repeat_region

/note="similar to EST BF369507 (NID:g11331532)"

repeat_region 7364..7661
 /rpt_family="Alu"
 repeat_region 8029..8048
 /rpt_family="(CAAAA)n"
 repeat_region 8379..8678
 /rpt_family="Alu"
 misc_feature 8560..8762
 /note="similar to EST AA699398 (NID:g2702592) z140a03.s1"
 repeat_region 8720..9031
 /rpt_family="Alu"
 misc_feature 9083..9560
 /note="similar to EST AA633603 (NID:g2556817) ae66b02.s1"
 repeat_region 9598..9899
 /rpt_family="Alu"
 misc_feature 9871..10098
 /note="similar to EST AA699398 (NID:g2702592) z140a03.s1"
 repeat_region 10746..11044
 /rpt_family="L1"
 repeat_region 11075..11101
 /rpt_family="AT-rich"
 repeat_region 11102..11372
 /rpt_family="Alu"
 repeat_region 11411..11465
 /rpt_family="GA-rich"
 repeat_region 11552..11644
 /rpt_family="L2"
 repeat_region 11685..12193
 /rpt_family="L1"
 repeat_region 12194..12486
 /rpt_family="Alu"
 repeat_region 12487..13003
 /rpt_family="L1"
 repeat_region 13004..13305
 /rpt_family="Alu"
 repeat_region 13306..13501
 /rpt_family="L1"
 repeat_region 13503..13892
 /rpt_family="L1"
 misc_feature 13893..14224
 /note="similar to EST BE061102 (NID:g8405752)"
 repeat_region 14244..14544
 /rpt_family="Alu"
 repeat_region 14546..14828
 /rpt_family="Alu"
 repeat_region 15091..15189
 /rpt_family="MIR"
 repeat_region 15672..16449
 /rpt_family="Achobo"
 repeat_region 16640..16844
 /rpt_family="Alu"
 repeat_region 16917..16959
 /rpt_family="Achobo"
 repeat_region 16960..17287
 /rpt_family="Alu"
 repeat_region 17288..17572
 /rpt_family="Achobo"
 misc_feature 17610..17938
 /note="similar to EST BF996427 (NID:g12402750)"
 repeat_region 17793..18068
 /rpt_family="Alu"
 repeat_region 18251..18537
 /rpt_family="Alu"
 misc_feature 18485..18884
 /note="similar to EST BE062347 (NID:g8406997)"
 misc_feature 18661..18809
 /note="similar to EST BG186544 (NID:g13708231)"
 repeat_region 18957..19160
 /rpt_family="L1"
 repeat_region 19314..19528
 /rpt_family="L1"
 repeat_region 19819..20135
 /rpt_family="Alu"
 repeat_region 21101..21269

```

repeat_region /rpt_family="L1"
21882..22203 /rpt_family="L1"
misc_feature 22429..22797 /note="similar to EST BF993478 (NID:g12399801)"
repeat_region 22717..23007 /rpt_family="L1"
misc_feature 22993..23191 /note="similar to EST BF767044 (NID:g12115035)"
repeat_region 23167..23471 /rpt_family="Alu"
misc_feature 23511..23930 /note="similar to EST AW105611 (NID:g6076346) xd49g02.xl"
repeat_region 25387..25677 /rpt_family="Alu"
repeat_region 25706..26013 /rpt_family="Alu"

```

```

Query Match 64.7%; Score 79.6; DB 9; Length 212827;
Best Local Similarity 83.1%; Pred. No. 3.6e-17;
Matches 103; Conservative 0; Mismatches 19; Indels 2; Gaps 1;

QY 1 CAATTGAAAGAGTTCTTCTAGTGGTGAAGGCCCAACACACTGTGTCTTGCAGTGAG 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 176040 CAATTTAAATAATCTCTGTAGTGGCTGAAGGTCCTCCAGCTGTATCTCGCCAGTGAG 175981

QY 61 TTAGGTTGTACAGACGGCTGTAGCACTAGCGC--TTGACAGAACCTCAGACGCCAAAG 118
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 175980 TTAAGTTGTACAGACATCTGTCAGCACTAGCAGATTACAGAACCTCAGACGCCAAAG 175921

QY 119 GTAC 122
| ||
Db 175920 GAAC 175917

```

```

RESULT 12
G31709
LOCUS G31709 262 bp mRNA linear STS 28-SEP-1998
DEFINITION SWSS2233 Eric D. Green Homo sapiens STS cDNA, sequence tagged site.
ACCESSION G31709
VERSION G31709.1 GI:3660607
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 262)
AUTHORS Bouffard,G.G., Iyer,L.M., Idol,J.R., Braden,V.V., Cunningham,A.F.,
Weintraub,L.A., Mohr-Tidwell,R.M., Peluso,D.C., Fulcon,R.S.,
Leckie,M.P. and Green,E.D.
TITLE A collection of 1814 human chromosome 7-specific STSs
JOURNAL Genome Res. 7 (1), 59-64 (1997)
MEDLINE 97189344
PUBMED 9037602
REFERENCE 2 (bases 1 to 262)
AUTHORS Green,E.D.
TITLE Human chromosome 7 STSs (1997)
JOURNAL Unpublished
COMMENT On Sep 29, 1998 this sequence version replaced gi:1916434.
Synonyms: AB4082
Contact: Eric D. Green
Genome Technology Branch
National Human Genome Research Institute/NIH
49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892
Tel: 3014020201
Fax: 3014024735
Email: egreen@nhgri.nih.gov
Primer A: CACGCTGATTCCTGCCAGT
Primer B: GGACACACGCCCTCTGTAGT
STS size: 123
PCR Profile:
Presoak: 0 degrees C for 0.00 minute(s)
Denaturation: 92 degrees C for 0.17 minute(s)

```

```

Annealing: 55 degrees C for 1.00 minute(s)
Polymerization: 72 degrees C for 1.00 minute(s)
PCR Cycles: 35
Thermal Cycler: PerkinElmer 9600

Protocol:
Template: 30-100 ng
Primer: each 1 uM
/dNTPs: each 200 uM
Taq Polymerase: 0.05 units/ul
Total Vol: 10 ul

Buffer:
MgCl2: 1.5 mM
KCl: 100 mM
Tris-HCl: 10 mM
NH4Cl: 5 mM
pH: 8.6

```

This STS has been incorporated into the NHGRI chromosome 7 physical map, but was developed by another investigator. See GenBank record: F03612 For additional information about the NHGRI chromosome 7 mapping project, see <http://www.nhgri.nih.gov/DIR/STB/CHR7>. Also see Genomics 11:548-64 (1991) [MUID=92128937].

```

FEATURES
source
    source          1..262
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /map="7"
                    /clone_lib="Eric D. Green"
                    134..256
                    134..153
    primer_bind    complement(237..256)
    primer_bind    88 a 55 c 53 g 64 t 2 others
BASE COUNT
ORIGIN

```

```

Query Match 63.1%; Score 77.6; DB 11; Length 262;
Best Local Similarity 81.5%; Pred. No. 1e-16;
Matches 101; Conservative 0; Mismatches 21; Indels 2; Gaps 1;

QY 1 CAATTGAAAGAGTTCTTCTAGTGGTGAAGGCCCAACACACTGTGTCTTGCAGTGAG 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 97 CAATTTAAATAATCTCTGTAGTGGCTGAAGGTCCTCCAGCTGTATCTCGCCAGTGAG 156

QY 61 TTAGGTTGTACAGACGGCTGTAGCACTAGCGC--TTGACAGAACCTCAGACGCCAAAG 118
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 157 TTAAGTTGTACAGACATCTGTCAGCACTAGCAGATTACAGAACCTCAGACGCCAAAG 216

QY 119 GTAC 122
| ||
Db 217 GAAC 220

```

```

RESULT 13
AC104663/c
LOCUS AC104663 137625 bp DNA linear PRI 10-FEB-2002
DEFINITION Homo sapiens chromosome 4 c-clone RP11-713M6, complete sequence.
ACCESSION AC104663 AC024963
VERSION AC104663.3 GI:18643727
KEYWORDS HYG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 137625)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 137625)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,

```


QY	1	CAATTGAAAAAATTTTCTTAGTGGTGCAGAAAGCCCAACACTGTGTCTTGGCAGTGAG	60
Db	46386	CAACTTAAAAAATTTTCTTAGTGGTGTGAAGGTCCTCCCTTGCTACTCTTGGCAGTGAG	46445
QY	61	TTAGGTGTGCACAAAGCGGTAGCACTAGGC--TTGCAGAACCTTCACAGACCCAAAG	118
Db	46446	TTAAGTGTGTAGAACAATCGTCAGCAGCAGCAGTGTACAGAACCTTCACAAACCCAAAG	46505
QY	119	GTAC 122	
Db	46506	GAAC 46509	
RESULT 15			
CNS01DS5			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	7	AAAAAAGTTTCTTAGTGGTGCAGAAAGCCCAACACTGTGTCTTGGCAGTGTAGT	66
Db	49673	AAAAAANTGTTCTAAGTGGTGTGAAGTCCAACTGTGTATTTTGCAGTGAGTTAAGT	49732
QY	67	TGTACAGAACGGCGTTAGCACTAGCGCT	94
Db	49713	TGTACAGAACAGCCAGGCAGCACTGGCT	49760

Sequence 159960 BP; 51462 A; 28399 C; 28309 G; 49987 T; 1803 other;

11